

ON CONTEMPORARY MORTALITY MODELS FOR ACTUARIAL USE II: PRINCIPLES

BY ANGUS S. MACDONALD[†] AND STEPHEN J. RICHARDS[‡]

ABSTRACT

We reprise some common statistical models for actuarial mortality analysis using grouped counts. We then discuss the benefits of building mortality models from the most elemental items. This has two facets. First, models are better based on the mortality of individuals, rather than groups. Second, models are better defined in continuous time, rather than over fixed intervals like a year. We show how survival probabilities at the ‘macro’ level arise at the ‘micro’ level from a series of Bernoulli trials over infinitesimally small time periods. Using a multiplicative representation of the mortality hazard rate, we show how counting processes naturally represent left-truncated and right-censored actuarial data, individual or age-grouped. Together these explain the ‘pseudo-Poisson’ behaviour of survival model likelihoods.

KEYWORDS

Bernoulli Trial, Counting Process, Likelihood, Product-integral, Pseudo-Poisson Model

CONTACT ADDRESSES

[†] Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK, and the Maxwell Institute for Mathematical Sciences, UK.

[‡] Longevity Ltd., 24a Ainslie Place, Edinburgh EH3 6AJ

Corresponding author: Angus Macdonald, A.S.Macdonald@hw.ac.uk.

If I dared, I would say we must have a theory — the word ‘theory’ is so much disliked by so many Englishmen, and is considered by them so ‘unpractical’, that I avoid it all I can; though I cannot see, myself, that it is very ‘practical’ to do things without knowing the theory of how to do them.

Wintringham & Blashford-Snell (1973)

1. INTRODUCTION

1.1 *In Search of a Continuous-time Model of Mortality*

Richards & Macdonald (2024) set out some practical benefits of using ‘continuous-time’ models of mortality. This expository paper asks what we mean by a ‘continuous-time’ model of mortality. As we seek an answer, in the theoretical basis of actuarial mortality modelling, we provide the language and notation to keep actuaries abreast of some fairly recent developments. The purpose of the paper is not to provide novel results, but to demonstrate how an actuary can see familiar objects in novel ways.

In fact, the idea of a ‘continuous-time’ model of mortality is not clear-cut or self-contained, and it leads us to consider two contrasts, which we may think of as modelling phenomena on a ‘micro’ scale versus phenomena on a ‘macro’ scale. These are:

- (a) very informally, the choice of infinitesimal time unit dx versus a discrete time unit, which we take to be a year; and
- (b) models based on individual lives versus models based on collectives of lives, including, inter alia, the collection of data based on individual lives versus collection of age-grouped data.

1.2 Inspiration from the Life Table

The life table is an obvious source of inspiration. Indeed, in the past some have viewed the whole subject as being the construction of life tables, see for example Batten (1978). A life table is a model of a cohort of identical and independent individuals, followed from some initial selection event at integer age $x_0 \geq 0$ (such as birth, with $x_0 = 0$) until mortality has extinguished the whole cohort, typically represented by the function l_x , interpreted as the expected number left alive at integer ages $x \geq x_0$. The two key features are:

- (a) the focus on the collective rather than the individual; and
- (b) the time unit of a year;

which are both ‘macro’ properties. If we model the number of deaths between integer ages x and $x + 1$ as a random variable D_x , then this formulation of the life table immediately suggests the binomial distribution as a model for D_x , see Section 2.3.

A slightly different view of the life table inspires a different model. Allow l_x to range over all real $x \geq x_0$, not just integer ages, and interpret the ratio l_{x+n}/l_x as the probability that an individual alive at age x survives to age $x + n$ ($x \geq x_0, n \geq 0$). This leads to further ideas, namely:

- (a) a model in which death is possible at any moment of time; and
- (b) the force of mortality or hazard rate (our preferred term) μ_x at age x as a measure of the instantaneous risk of death;

which are both ‘micro’ properties. However, observation is still ‘macro’, of the collective rather than of the individual. This setup suggests a Poisson model for D_x , see Section 2.4.

1.3 Models Based on Individuals: The Pseudo-Poisson Model

More recent introductions to the subject begin with the definition of the future lifetime of a person age x as a non-negative random variable, denoted by T_x . For brevity and completeness, we compress the definitions of related quantities into Table 1, see Dickson et al. (2020) or Macdonald et al. (2018) for details. Of course, the actuarial symbols ${}_tq_x, {}_tp_x$ and μ_x would be defined, based on the life table, in the process of obtaining the binomial and Poisson models in Section 1.2, but the point is that they are now defined by their rôles in the distribution of T_x .

If we define T_0^i to be the random lifetime of the i th individual under observation, this model focusses attention on:

- (a) the individual rather than the collective; and
- (b) events happening instantaneously, meaning during a short time period h as we let $h \rightarrow 0^+$;

Table 1: Definitions of quantities based on T_x , the random future lifetime at age x . The consistency condition assumes that $x_0 = 0$, and ensures that calculations based on the distribution of T_x will never contradict calculations based on the distribution of T_y ($y \neq x$).

Quantity	Definition
Distribution function:	$P[T_x \leq t] = {}_tq_x$
Survival function:	$P[T_x > t] = {}_tp_x$
Hazard rate :	$\lim_{h \rightarrow 0^+} \frac{{}_hq_x}{h} = \mu_x$
Density function:	$\frac{d}{dx} {}_tq_x = {}_tp_x \mu_{x+t}$
Consistency condition:	$P[T_x \leq t] = P[T_0 \leq x + t \mid T_0 > x]$

which are both ‘micro’ properties. The most important idea is expressed in the heuristic:

$$P[\text{Dead by age } x + h \mid \text{Alive at age } x] = {}_hq_x \approx h \mu_x \quad (\text{for small } h). \quad (1)$$

Knowing the density function of each T_x^i (Table 1), we can write down the probability of any observations, hence a likelihood, and that leads to the following explanation of why the Poisson model of Section 2.4 works so well. Suppose we assume a constant hazard rate at each age, we observe M individuals and there are D deaths. Then:

- (a) the model based on individual random lifetimes gives us, in principle, an exact probability of observing D deaths; while
- (b) the Poisson model of Section 2.4 gives us only an approximate probability of observing D deaths (it must do since $D \leq M$ but $P[D > M] > 0$ under a Poisson model); however
- (c) both models give us exactly the same likelihood of observing D deaths (up to irrelevant factors).

It follows that inference based on the likelihood will be identical under both models. This leads us to call the model based on individual lifetimes (and a constant hazard) the pseudo-Poisson model (Section 3.7). We continue to seek the proper foundations of a mortality model at the ‘micro’ level in the model of individual lifetimes.

1.4 Dynamic Life History Models I: Truncation and Censoring

The individual life-history model lets us write down exact probabilities of observed events, if we know the hazard rates. It also lets us deal with incomplete observation, in particular:

- (a) left-truncation: an individual enters observation having already survived to some age $x_a > 0$; and
- (b) right-censoring: the individual leaves observation while still alive, so we observe only that $T_{x_a} > x_b$ for some age $x_b > x_a$.

A neat device allows us to avoid the complication of keeping track of ages x_a and x_b when writing expressions such as likelihoods. Define a process:

$$Y^i(x) = I_{\{\textit{i}th \textit{individual alive and under observation at age } x^-\}} \quad (2)$$

(age x^- means ‘just before age x ’ and is a technicality). The ‘under observation’ condition takes care of left-truncation and right-censoring. Then, for example, the integrated hazard rate over the time spent under observation by the i th individual (an important quantity in many calculations), can be written:

$$\int_{x_a}^{x_b} \mu_x dx = \int_0^\infty Y^i(x) \mu_x dx. \quad (3)$$

We see that the process $Y^i(x) \mu_x$ acts as a dynamic or stochastic hazard rate tailored to the i th individual, and greatly simplifies expressions involving integrals, since all integrals can now be now taken over $(0, \infty]$ (Section 4.6).

1.5 Dynamic Life History Models II: Back to Bernoulli

In a utilitarian sense the job was finished with Section 1.3, but the heuristic (1) suggests more to come. For, if ${}_h q_x \approx h \mu_x$ then ${}_h p_x \approx 1 - h \mu_x$, and if we let δ_x be an indicator, equal to 1 if death occurs at age x , and 0 otherwise, then what is observed ‘during’ time h is the outcome of a Bernoulli trial with parameter $h \mu_x$ and probability:

$$(1 - h \mu_x)^{1-\delta_x} (h \mu_x)^{\delta_x}. \quad (4)$$

We would like to take all such consecutive ‘instantaneous’ Bernoulli trials while the individual is alive and under observation, and multiply their probabilities (4) together. In all of probability theory, there is nothing simpler than a Bernoulli trial, so we really would have reduced a probability in a mortality model to its constituent ‘atoms’; the ultimate ‘micro’ level. That is what we describe in Section 4. To do so we introduce two ideas, which give us the notation needed to write down a product of Bernoulli probabilities like (4) in a rigorous way.

- (a) Counting processes: A counting process $N^i(x)$ starts at $N^i(0) = 0$ and jumps to 1 at time T_0^i if the i th individual is then under observation. Then its increment $dN^i(x)$ indicates an observed death, and is a rigorous version of the informal δ_x in (4). Between them, $N^i(x)$ and $Y^i(x)$ let us write the Bernoulli trial probability (4) formally as:

$$(1 - Y^i(x) \mu_x dx)^{1-dN^i(x)} (Y^i(x) \mu_x dx)^{dN^i(x)} \quad (5)$$

and this allows for left-truncation and right-censoring.

- (b) Product-integral: The product-integral is the device that lets us multiply all the infinitesimal Bernoulli trial probabilities. We defer further description to Section 4.2 and Appendix 2 and just give the final form of the likelihood contributed by the i th individual, denoted by L_i :

$$L_i = \prod_{x \in (0, \infty]} (1 - Y^i(x) \mu_x dx)^{1-dN^i(x)} (Y^i(x) \mu_x dx)^{dN^i(x)}. \quad (6)$$

The product-integral is identified by a product over all values of an interval ($x \in (0, \infty]$ here) and the presence of the variable of integration (dx here) in the integrand.

Therefore, we have a mortality model with the following properties:

- (a) It is irreducible, in the sense that it is composed of consecutive (infinitesimal) Bernoulli trials.
- (b) It is based on behaviour at the ‘micro’ time scale.
- (c) It is based on individual lives.
- (d) Aggregated, over time and over individuals, it explains the Poisson-like nature of likelihoods, therefore estimation based on the collective at the ‘macro’ time scale.
- (e) It allows for left-truncation and right-censoring.
- (f) It is easily extended to multiple-decrement and multiple-state models.

We started out by trying to pin down what we meant by the vague term ‘continuous-time mortality model’. Now we have an answer, although our endpoint is just the starting point for the modern statistical study of survival models (Section 4.10), see Andersen et al. (1993).

1.6 Plan of this Paper

We start in Section 2 with Forfar et al. (1988), a definitive account of graduation using binomial and Poisson models, which we call mortality models at the ‘macro’ scale. Then in Section 3 we turn to models at the ‘micro’ scale based on individual lifetimes, and find the origins of Poisson-like behaviour at the ‘macro’ scale. In Section 4 we bring together models of individual lifetimes and models based on behaviour over small intervals h as $h \rightarrow 0^+$, and find that all probabilities in a mortality model arise as a product of consecutive (infinitesimal) Bernoulli trials. Section 5 concludes.

2. BINOMIAL AND POISSON MODELS

2.1 Forfar et al. (1988)

In a landmark paper, Forfar et al. (1988) gave comprehensive accounts of two models for survival data, namely the binomial and Poisson models. These defined: (a) the random variable D_x , to be the number of deaths observed at age x ; and (b) a suitable measure of exposure to risk at age x , assumed to be non-random, that we will call V_x . Then the occurrence-exposure rate D_x/V_x was shown to be an estimate of the model parameter, a mortality rate \hat{q} in the binomial model, and a hazard rate $\hat{\mu}$ in the Poisson model¹.

Forfar et al. (1988) helpfully located the old subject of parametric graduation in a modern statistical setting, including model specification, likelihood, score function and information, covariance matrix, model selection and parametric bootstrapping. The treatment was heavily influenced by the authors’ work for the Continuous Mortality Investigation Bureau (the CMIB, now CMI) particularly in respect of data collection. The advance it represented may be gauged by comparison with contemporary texts such as Batten (1978) and Benjamin & Pollard (1980).

¹A third model was offered, identical to the Poisson model except that the occurrence-exposure rate was taken to be an estimate \hat{m} of the life table quantity m_x , the central rate of mortality (see Neill (1977) for example). Since m_x does not arise naturally as the parameter in any well-defined statistical model we will not pursue this further.

Both binomial and Poisson models are rooted in simple thought-experiments, require no statistics beyond a first course in data analysis, and can, with qualifications, be implemented in standard statistical packages such as R (R Core Team 2021). This gives them considerable staying power.

2.2 The Rate Interval and Δ_k Notation

The rate interval is an interval of age (or calendar time) on which an individual is assigned a given age label. It is the means of assigning an age label to an individual exposed to the risk of death and at the time of death. Note that rate intervals are only needed with age-grouped data, not models based on individual lives. They are treated in detail in texts such as Benjamin & Pollard (1980). We assume that the rate interval, when we need one, is the year of age $(x, x + 1]$ defined by ‘age last birthday’. The CMI, for another example, use the year of age $(x - 1/2, x + 1/2]$ defined by ‘age nearest birthday’.

We assume that the data are covered by K rate intervals and that in the abstract these may be denoted by Δ_k ($k = 1, 2, \dots, K$); and that a sum over all rate intervals may be denoted by \sum_k , and a product likewise by \prod_k .

2.3 Binomial Models

The binomial model is based on the following thought-experiment: take E_x lives at the start of a year, all alive at age x and assumed to be ‘statistically independent’ in respect of their mortality risk. Then E_x is the measure of exposure referred to as V_x in Section 2.1, usually here called the initial exposed-to-risk. Define D_x to be the number who are dead at the end of the year, and q_x to be the probability that a life alive at age x dies not later than age $x + 1$. Then the following are easily shown.

- (a) D_x has a binomial(E_x, q_x) distribution, with first two moments $E[D_x] = E_x q_x$ and $\text{Var}[D_x] = E_x q_x (1 - q_x)$.
- (b) As a function of parameter q_x , the data (D_x, E_x) has likelihood function:

$$\begin{aligned} L(q_x) &= \binom{E_x}{D_x} q_x^{D_x} (1 - q_x)^{E_x - D_x} \\ &\propto q_x^{D_x} (1 - q_x)^{E_x - D_x}, \end{aligned} \quad (7)$$

leading to the maximum likelihood estimate (MLE) $\hat{q}_x = D_x/E_x$ which is asymptotically unbiased ($E[\hat{q}_x] = q_x$) with variance $\text{Var}[\hat{q}_x] = q_x(1 - q_x)/E_x$.

- (c) The estimate \hat{q}_x is an estimate of q_x , that is, the function value at start of the rate interval $(x, x + 1]$.

D_x can be viewed as the number of successes out of E_x independent Bernoulli trials, each with probability of success (death) equal to q_x . The idea of the Bernoulli trial as the fundamental ‘atom’ of mortality risk appears again in Section 4.5.

2.4 Poisson Models

The Poisson model depends on a different thought-experiment. An unspecified number of individuals is observed, alive during the relevant rate interval of age $(x, x + 1]$, such that the total time alive and under observation is a non-random quantity E_x^c . Now E_x^c is

the measure of exposure referred to as V_x in Section 2.1, usually here called the central exposed-to-risk. A constant force of mortality μ is assumed at all ages in the rate interval $(x, x + 1]$. Define D_x to be the number of observed deaths. Then the following can be shown.

- (a) D_x has a Poisson(μE_x^c) distribution with first two moments $E[D_x] = \text{Var}[D_x] = \mu E_x^c$.
- (b) As a function of parameter μ , the data (D_x, E_x^c) has likelihood function:

$$\begin{aligned} L(\mu) &= (\mu E_x^c)^{D_x} \exp(-\mu E_x^c) / D_x! \\ &\propto \exp(-\mu E_x^c) \mu^{D_x}, \end{aligned} \tag{8}$$

leading to the MLE $\hat{\mu} = D_x / E_x^c$ which is asymptotically unbiased ($E[\hat{\mu}] = \mu_x$) with variance $\text{Var}[\hat{\mu}] = \mu / E_x^c$.

- (c) Assuming a relatively even distribution of exposure over the rate interval $(x, x + 1]$, the MLE $\hat{\mu}$ estimates $\mu_{x+1/2}$.

2.5 Terminology

Binomial and Poisson models may be described in different ways. The binomial model admits of no conceivable time other than its own time unit; it is unambiguously a discrete-time model. It may also be called a q-type model in honour of its conventional parameter. The Poisson model is a candidate for a continuous-time model, although it turns out to be an extreme representative of a whole class, see Sections 3 et seq.. It may also be called a μ -type model in honour of its conventional parameter. See Richards & Macdonald (2024) for both terminologies.

2.6 Assessment of the Binomial and Poisson Models

2.6.1 Feasibility of the Thought-Experiment: Binomial Model

To carry out the binomial thought-experiment we would need a homogeneous sample of E_x individuals age x , observed to be alive or dead at age $x + 1$. This contrasts with observation of (say) members of a pension scheme or life office policyholders. Real data often includes exits for reasons other than death and not under the modeller's control, see (Richards & Macdonald 2024, Section 3 and Appendix) for examples. The requirements of the binomial experiment will not be met by: (a) individuals entering observation between ages x and $x + 1$; and (b) individuals leaving observation between ages x and $x + 1$, for reasons other than death.

Thus we are led to ask, what is the probability of surviving over any fraction of the rate interval? For example, an individual joining at age $x - 1/2$ and surviving to age x requires the calculation of ${}_{1/2}p_{x-1/2}$. The binomial model gives no satisfactory answer. Strictly, the question lies outside the bounds of the model. Even if we could implement the thought-experiment, the model posits only the number of lives observed at the start and end of the rate interval.

Nevertheless, an answer may be demanded, because individuals can and do join or leave an investigation in the middle of the rate interval, see (Richards & Macdonald 2024, Section 3) for numerous examples in practice. The analyst is obliged to make some

assumption about mortality between ages x and $x + 1$, for which the binomial model gives no guidance. Three popular assumptions have been:

- (a) a uniform distribution of deaths;
- (b) the Balducci hypothesis;
- (c) a constant hazard rate.

See Macdonald (1996) or Richards & Macdonald (2024) for a discussion of these. Here we just remark that (c), a constant hazard rate, is mathematically the simplest, fully consistent with the Poisson model, and also consistent with modelling individual lifetimes as in Section 3.

2.6.2 Feasibility of the Thought-Experiment: Poisson Model

The Poisson thought-experiment is not troubled by fractions of the rate interval. Since the hazard rate is assumed to be a constant μ during the rate interval $(x, x + 1]$, the probability of dying during the sub-interval $(x + a, x + b]$ (given alive at age $x + a$, for $0 \leq a < b \leq 1$) is $1 - \exp(-\mu(b - a))$.

The Poisson thought-experiment is not met in practice, however, for different reasons. The distribution of D_x is Poisson only if the exposed-to-risk E_x^c is non-random, for example pre-determined. This is not the case if: (a) the population being sampled is finite, with known maximum size M individuals, say, because then $D_x \leq M$, but $P[D_x > M] > 0$ under any Poisson distribution; or (b) the exposure times of the individuals in the sampled population are not known in advance, because then E_x^c is random. Moreover, D_x is usually a component of the bivariate random variable (D_x, E_x^c) . In such cases we call D_x pseudo-Poisson, see Section 3. For estimation purposes, however, it behaves as a true Poisson random variable would, see Section 3.7.

2.6.3 Occurrence-exposure Rates, Age-grouped Data and Graduation

The estimates $\hat{q}_x = D_x/E_x$ and $\hat{\mu}_x = D_x/E_x^c$ are examples of occurrence-exposure rates. Both they and their sample variances (Sections 2.3 and 2.4) require only the age-grouped totals D_x and E_x or E_x^c to be reported to the analyst, rather than data on each individual. Such totals may easily be extracted from ordinary data files used in the business; they greatly reduce the volume of data required (which used to matter a lot); and they reduce the risk of accidentally breaching data-protection rules (which matters now). On the other hand they do not allow the level of checking and cleaning of the data that is possible with individual data (Macdonald et al. 2018, Chapter 2).

If age-grouped data are prepared by someone other than the analyst, the modelling is wholly dependent on the thoroughness and diligence of the source provider. This is a material concern for risk-transfer transactions, such as reinsurance, bulk annuities and portfolio transfers. If a model is to be used to price a risk transfer, the analyst should always insist on individual records, regardless of whether the intent is to use models based on individuals or age-grouped counts.

Occurrence-exposure rates \hat{q}_x or $\hat{\mu}_x$ are normally smoothed or graduated for practical use. For this purpose a likelihood may be calculated as the product of the likelihoods for each rate interval, using age-grouped data. Other, non-likelihood methods may also be used (Forfar et al. 1988).

2.7 Generalized Linear Models (GLMs)

To the list of properties in Sections 2.3 and 2.4 we could have added “(d) Leads to a simple Generalized Linear Model (GLM) for graduating age-grouped mortality data.”

GLMs were introduced by Nelder & Wedderburn (1972), and contain three elements: (a) a random component, Y_x ; (b) a systematic component, η_x ; and (c) a link function, g . A GLM connects the expectation of Y_x to η_x via g as follows:

$$\eta_x = g(\mathbb{E}[Y_x]). \quad (9)$$

The component η_x is the linear predictor; in mortality work it is a linear function of age, x , and a corresponding covariate vector, \mathbf{z}_x . Let $\boldsymbol{\theta}$ be the vector of parameters to estimate, and let \mathbf{X} be the corresponding model matrix. Each observation Y_x has a corresponding row in \mathbf{X} . For a binomial GLM we have:

$$Y_x = \frac{D_x}{E_x}, \quad \eta_x = \mathbf{X}\boldsymbol{\theta}[x,]. \quad (10)$$

For a Poisson GLM with the link function $g(x) = \log(x)$ we have:

$$Y_x = D_x, \quad \eta_x = \mathbf{X}\boldsymbol{\theta}[x,] + \log(E_x^c). \quad (11)$$

where $[x,]$ selects the row for the observation corresponding to age x .

The link function, g , is chosen by the analyst. The canonical link for the binomial GLM is the logit, but other link functions can be used, such as the probit link. The canonical link for the Poisson GLM is the logarithm, but other link functions have been used for mortality work, such as the logit link; see (Currie 2016, Appendix 1) for implementation details of the logit link for Poisson GLMs.

GLMs have a link with ‘classical’ actuarial modelling since one of the simplest choices of fitted $\hat{\eta}_x$ is a Gompertz function, but this does not extend to other members of the Gompertz-Makeham family (see Forfar et al. (1988)).

GLM’s are popular because they are flexible, have nice statistical properties and are linear in the covariates. The binomial and Poisson error structures arise naturally for ‘count’ data, and age-grouped deaths are examples of ‘counts’ so these GLMs are, in a sense, natural candidates as mortality models. However, linear dependence on covariates, and the canonical link functions associated with the exponential family, are restrictive, and for large experiences we will often find better-fitting models that are not GLMs (see, for example, the range of models included in Cairns et al. (2009)). In addition GLMs bring us no closer to any foundational concept of a ‘mechanism’ generating mortality data, so we do not consider them further.

3. MODELLING INDIVIDUAL LIFETIMES: THE PSEUDO-POISSON MODEL

3.1 Observation of an Individual

Suppose the i th individual is observed from age x_i until age y_i for total time $v_i = y_i - x_i$. Denote the interval $(x_i, y_i]$ by Δ_i . Observation ends because of either death or right-censoring at age y_i . Define the indicator:

$$d_i = \begin{cases} 1 & \text{if } i\text{th individual died at age } y_i \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

Then the random variable observed is the bivariate (d_i, v_i) , and the total contribution to the likelihood of these observations, denoted by L_i , is:

$$\begin{aligned} L_i &= v_i p_{x_i} \mu_{x_i+v_i}^{d_i} \\ &= \exp\left(-\int_0^{v_i} \mu_{x_i+s} ds\right) \mu_{x_i+v_i}^{d_i} \\ &= \exp\left(-\int_{\Delta_i} \mu_s ds\right) \mu_{y_i}^{d_i} \end{aligned} \quad (13)$$

see Table 1, or (Macdonald et al. 2018, Chapter 5).

3.2 Age Intervals and Δ_i Notation

The definition of the interval Δ_i depends on the observational plan and the method of investigation. The important point is that it is time under observation of a single individual, the i th of M individuals. Some examples are the following.

- (a) The interval may be the entire period for which the i th individual was observed, potentially spanning many years.
- (b) The interval may be that part of a rate interval (for example, year of age) for which the i th individual was under observation.
- (c) The interval may be an interval of age on which the hazard rate is assumed to be constant.

Therefore the likelihood (13) based on Δ_i may constitute the whole of the i th individual's contribution to the total likelihood, or only part of it. We will call a contribution to a likelihood of the form (13) a survival model likelihood, whether it forms all or part of the i th individual's contribution, and whether or not hazard rates are assumed to be piecewise-constant.

3.3 Multiplication and Factorization of Survival Model Likelihoods

Survival model likelihoods in respect of the same individual over contiguous intervals, multiplied together, give another survival model likelihood. In reverse, a survival model likelihood may be factorized into as many factors of like kind as we please. To see the multiplicative property, suppose the i th individual is observed on the contiguous intervals $\Delta_i^1 = (x_i, z_i]$ and $\Delta_i^2 = (z_i, y_i]$, and that indicators of death $d_i^{(1)}$ at time z_i and $d_i^{(2)}$ at time y_i are defined analogously to d_i above. Then if $\Delta_i = (x_i, y_i]$ as before:

$$\begin{aligned} \exp\left(-\int_{\Delta_i} \mu_s ds\right) \mu_{y_i}^{d_i} &= \exp\left(-\int_{\Delta_i^1 \cup \Delta_i^2} \mu_s ds\right) \mu_{y_i}^{d_i} \\ &= \exp\left(-\int_{\Delta_i^1} \mu_s ds\right) \mu_{z_i}^{d_i^{(1)}} \exp\left(-\int_{\Delta_i^2} \mu_s ds\right) \mu_{y_i}^{d_i^{(2)}} \end{aligned} \quad (14)$$

since necessarily $d_i^{(1)} = 0$ and $d_i^{(2)} = d_i$ if events on Δ_2 are not trivially null. Whether we regard this as factorizing a likelihood on Δ_i , or multiplying two likelihoods on Δ_i^1 and Δ_i^2 , does not matter for our purposes.

3.4 Rate Intervals, Piecewise-constant Hazards and Age-grouped Data

Recall from Section 2.2 that a rate interval is denoted by Δ_k . Here let Δ_k be the rate interval from integer age k to age $k + 1$, that is, $\Delta_k = (k, k + 1]$. Age-grouped data may then be denoted by total deaths d_k and total person-years exposure E_k^c falling within rate interval Δ_k .

It is instructive to group data on individual lives to reproduce age-grouped data, and to compare the resulting likelihoods. This is aided by Table 2, which shows the contributions to likelihoods of individual data for two individuals, treated three ways. The i th individual is observed from age 47 until right-censored at age 50. The j th individual is observed from age 47.6 until dying at age 49.3. We list contributions to the likelihood under three combinations of observational plan and model:

- (a) Rate intervals Δ_k , and a constant hazard rate on each rate interval, denoted by μ_k^* .
- (b) Rate intervals Δ_k , and a smooth hazard rate parametrized by θ , denoted by μ_x^θ (for example, a Gompertz-Makeham function).
- (c) Observation of complete lifetimes on age interval Δ_i , and a smooth hazard rate parametrized by θ , also denoted by μ_x^θ .

The contributions are shown in Table 2. In obvious notation, we may denote the contributions to the likelihoods under (a), (b) and (c) above by $L_{i,k}^*$, $L_{i,k}^\theta$ and $L_i^\theta = \prod_k L_{i,k}^\theta$ respectively. Likewise, collecting all contributions to rate interval Δ_k in columns (a) and (b), we may define the total likelihood contributed by Δ_k by $L_k^* = \prod_i L_{i,k}^*$ and $L_k = \prod_i L_{i,k}$ respectively. This leads to the following observations.

- (a) It is obvious from columns (b) and (c) that for any individual, the likelihood over the complete lifetime is the product of the likelihoods over each rate interval, see Section 3.3. In fact we have incorporated this in the notation, $L_i^\theta = \prod_k L_{i,k}^\theta$. It makes no difference if we split the the individual lives data and present them by rate interval. But this does not lead to any simplification, and age-grouped totals d_k and E_k^c play no part, because of the smooth hazard rate in the integrands.
- (b) Each entry in column (a) can be regarded as approximating its partner in column (b). For example in the third and sixth lines, we approximate μ_{49+s}^θ by μ_{49}^* for $0 \leq s < 1$; we show the sixth line below:

$$\exp\left(-\int_{0.0}^{0.3} \mu_{49+s}^\theta ds\right) \mu_{49.3}^\theta \approx \exp\left(-\int_{0.0}^{0.3} \mu_{49}^* ds\right) \mu_{49}^* = \exp(-0.3\mu_{49}^*) \mu_{49}^*. \quad (15)$$

Collecting together all such terms in μ_{49}^* we get the total likelihood:

$$L_{49}^*(\mu_{49}^*) = \exp(-\mu_{49}^* E_{49}^c) (\mu_{49}^*)^{d_{49}} \quad (16)$$

Table 2: Contributions to likelihoods of the i th individual, under observation from age 47 until right-censored at age 50, and the j th individual, under observation from age 47.6 until death at age 49.3, under three observational plans and assumptions: (a) annual rate interval Δ_k , piecewise-constant hazard rates; (b) annual rate interval Δ_k , smooth hazard rate parametrized by θ ; and (c) observation of complete lifetime age interval Δ_i , smooth hazard rate parametrized by θ .

Observational Plan and Likelihood:

Individual	Rate Interval	Rate Interval Δ_k	Rate Interval Δ_k	Rate Interval Δ_k	Complete Lifetime Δ_i
	Interval	Likelihood $L_k^*(\mu_k)$	Likelihood $L_{i,k}^\theta$	Likelihood $L_{i,k}^\theta$	Likelihood $L_i^\theta = \prod_k L_{i,k}^\theta$
i	$\left\{ \begin{array}{l} \Delta_{47} \\ \Delta_{48} \\ \Delta_{49} \end{array} \right\}$	(a)	$\exp(-1.0 \mu_{47}^*)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{47+s}^\theta ds\right)$	(c)
		$\exp(-1.0 \mu_{48}^*)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{48+s}^\theta ds\right)$	$\left. \right\} \exp\left(-\int_{47.0}^{50.0} \mu_s^\theta ds\right)$	
		$\exp(-1.0 \mu_{49}^*)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{49+s}^\theta ds\right)$		
	$\left\{ \begin{array}{l} \Delta_{47} \\ \Delta_{48} \\ \Delta_{49} \end{array} \right\}$	(b)	$\exp\left(-\int_{0.6}^{1.0} \mu_{47+s}^\theta ds\right)$		
		$\exp(-1.0 \mu_{48}^*)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{48+s}^\theta ds\right)$		
		$\exp(-0.3 \mu_{49}^*) \mu_{49}$	$\exp\left(-\int_{0.0}^{0.3} \mu_{49+s}^\theta ds\right) \mu_{49.3}^\theta$		

in which the age-grouped totals do appear. Comparing equations (16) and (8), we see that the former is functionally identical to the likelihood from the Poisson model, and yet no assumption about Poisson random variables or distributions has been made in this section. In other words, the Poisson-like nature of the likelihood arises from the fundamental nature of modelling individual lifetimes.

3.5 *Individual versus Age-Grouped Data for Multiple Lives*

Table 2 illustrates how the individual lifetime model is related to age-grouped data, based on rate intervals, exactly without using any approximations. Indeed, columns (b) and (c) show that labelling the data by individual i or by rate interval k is merely a rearrangement. Specifically, the i th individual contributes $L_{i,k}^\theta$, possibly null, to the likelihood in rate interval k (fourth column). The outer form of the total likelihood, denoted by L^θ , then depends simply on the order in which we take products, as the following identities show:

$$L^\theta = \prod_i L_i^\theta = \prod_i \prod_k L_{i,k}^\theta = \prod_k \prod_i L_{i,k}^\theta = \prod_k L_k^\theta = L^\theta. \quad (17)$$

This informal statement based on Table 2 ('proof-by-example') of course needs to be demonstrated properly. Doing so with the notation to hand is surprisingly detailed, though elementary, and is delegated to Appendix 1. A much simpler proof will be shown when the notation of Section 4 is available (Section 4.7).

3.6 *The Rôle of Occurrence-exposure Rates*

We may arrive at the likelihood based on the age-grouped data (d_k, E_k^c) in two different ways.

- (a) We could use the Poisson model with parameter $\mu_k^* E_k^c$ (Section 2.4) for rate interval Δ_k .
- (b) Within the individual lives model, we could assume that the hazard rate is piecewise-constant with value μ_k^* on rate interval Δ_k . This means assuming that the parameter θ is the vector of hazard rates μ_k^* .

In either case, on rate interval Δ_k , we have a single parameter, which we denote by μ_k^* , and a likelihood that we denote by $L_k^*(\mu_k^*) = \exp(-\mu_k^* E_k^c) (\mu_k^*)^{d_k}$. In total we have a K -parameter model with likelihood $\prod_k L_k^*(\mu_k^*)$, from which the parameters are estimated independently by the occurrence-exposure rates d_k/E_k^c , which we denote by $\hat{\mu}_k^*$. That is as far as the probabilistic model takes us.

In traditional actuarial terminology, the $\hat{\mu}_k^*$ are 'crude' rates which require to be smoothed or graduated, using no more than the available age-grouped data (Benjamin & Pollard 1980). A convenient way of doing so is to use: (a) the likelihood function $\prod_k L_k^*(\mu_k^*)$; (b) a parametric function μ_x^θ for the hazard rate, of much lower dimension than K ; and (c) to connect the two with an assumption that $\hat{\mu}_k^*$ estimates $\mu_{x_k}^\theta$, for some $x_k \in \Delta_k$, for example $\hat{\mu}_k^*$ estimates $\mu_{k+1/2}^\theta$. Note that this smoothing procedure is not part of the probabilistic model, despite its use of the likelihood function. Forfar et al. (1988) show that it is approximately equivalent to the much older minimum- χ^2 method.

In either case, again, the age-grouped quantities approximate exact quantities as follows:

$$(\mu_k^*)^{d_k} \approx \prod_{d_{i,k}=1} \mu_{y_i}^\theta \quad \text{and} \quad \mu_k^* E_k^c \approx \sum_i \int_{\Delta_{i,k}} \mu_s^\theta ds \quad (18)$$

where $d_{i,k}$ is the number of deaths (0 or 1) befalling the i th individual in rate interval Δ_k , and $\Delta_{i,k} = \Delta_i \cap \Delta_k$ (possibly \emptyset). Therefore, inference based upon age-grouped data is close, but not identical, to inference based upon the individual lives data.

The crude hazard rates $\hat{\mu}_k^*$, or more accurately the expected deaths based upon them, $\hat{\mu}_k^* E_k^c$, may be used in forming statistics such as deviances, used in testing the fit of a graduation (Forfar et al. 1988, Benjamin & Pollard 1980, Macdonald et al. 2018).

3.7 Pseudo-Poisson Models

The various likelihoods that appear in this section, see Table 2, are all Poisson-like, and if a piecewise-constant hazard rate is assumed, indistinguishable from a true Poisson likelihood, an observation that goes back to the earliest work on inference in Markov models, see for example Sverdrup (1965), Waters (1984). However, there are no Poisson random variables. In the likelihood (13): (a) d_i is either 0 or 1, and does not range over the non-negative integers; (b) v_i is random, not deterministic; and (c) d_i and v_i are not independent, the random variable is the bivariate (d_i, v_i) .

Many authors suppose, as we did in Section 2.4, that the number of deaths in some model has a Poisson distribution, but without ensuring, as we did, that the exposure times would be non-random; more often the observation of random death times ensures the opposite. This conceptual error is almost always immaterial for inference, precisely because the Poisson likelihood is of the correct form for a survival model, although the survival model is not Poisson. Where it matters is in misdirecting us when we come to extend the survival model, including allowing for: (a) truncation and censoring; (b) more complicated life histories, including multiple decrements; (c) calculating residuals when the expected number of deaths is small; and (d) statistics for multiple lives, see Section 3.5 and Appendix 1.

We suggest it would be clearer and less confusing if the term pseudo-Poisson was adopted, to describe the great majority of models for death counts that appear in the literature.

3.8 Covariates

Covariates may be introduced by defining a vector \mathbf{z}^i of covariates for the i th individual and letting the hazard rate be a function $\mu(x, \mathbf{z}^i)$ of age and covariates. A common way to introduce such a dependency is to define a vector $\boldsymbol{\beta}$ of regression coefficients such that the hazard rate is a function $\mu(x, \boldsymbol{\beta}^T \mathbf{z}^i)$ of age and a linear combination of the covariates. Further simplification is achieved if the hazard rate factorizes as $\mu(x, \mathbf{z}^i) = \mu_x \times g(\mathbf{z}^i)$, the product of an age dependent hazard rate μ_x (called the baseline hazard) and some function g of the covariates; then the hazard rates of any two individuals of the same age are always in the same proportion, called proportional hazards. Finally, the most common choice of g is an exponential function of a linear combination of the covariates, $\mu(x, \mathbf{z}^i) = \mu_x \times \exp(\boldsymbol{\beta}^T \mathbf{z}^i)$, which has proportional and non-negative hazards as well as a

Table 3: Three stages in adding structure to a hazard rate that is a function $\mu(x, \mathbf{z}^i)$ of age x and a vector \mathbf{z}^i of covariates for the i th individual. Each stage is increasingly restrictive, from the most flexible model in Stage 0 to the most restrictive in Stage 3.

Stage	Form of $\mu(x, \mathbf{z}^i)$	Description
0	$\mu(x, \mathbf{z}^i)$	General function of x and \mathbf{z}^i
1	$\mu(x, \boldsymbol{\beta}^T \mathbf{z}^i)$	Function of age and linear combination of covariates
2	$\mu_x \times g(\mathbf{z}^i)$	Proportional hazards
3	$\mu_x \times \exp(\boldsymbol{\beta}^T \mathbf{z}^i)$	Basis of the Cox model (Cox 1972)

log-linear dependence on covariates. These steps in adding structure to the hazard rate are summarized in Table 3.

The last hazard structure in Table 3 is popular in medical statistics, where it is known as the Cox model, because the baseline hazard can be ignored and only the regression coefficients need to be estimated, by the procedure known as partial likelihood (see Andersen et al. (1993)). However, actuaries usually wish to estimate the whole model, baseline hazard included, whatever the form of the hazard rate. Then the full likelihood (40) from Appendix 1 becomes:

$$L = \prod_{i=1}^M L_i = \prod_{i=1}^M \prod_{k=1}^K \prod_{\Delta_j \subseteq \Delta_k} \left[\exp \left(- \int_{\Delta_j} \mu(s, \mathbf{z}^i) ds \right) \right]^{e_{i,j}} \mu(w_j, \mathbf{z}^i)^{d_{i,j}}. \quad (19)$$

Clearly any of the hazard rates in Table 3 may be substituted into the likelihood (19). However inspection of the innermost elements of (19), integrals over intervals $\Delta_j \subseteq \Delta_k$, shows that, even if the hazard rate factorizes as in Stages 2 and 3 of Table 3, these factors cannot be collected together to form likelihoods L_k over rate intervals. See also the written comments by A. D. Wilkie in the discussion of Richards (2008).

4. DYNAMIC LIFE HISTORY MODELS

4.1 The Anatomy of a Survival Probability

We begin with a closer examination of the multiplicative property of survival probabilities, usually expressed as:

$${}_{s+t}p_x = {}_t p_x {}_s p_{x+t} = {}_s p_x {}_t p_{x+s}. \quad (20)$$

We can apply this repeatedly to factorize ${}_t p_x$, with n a positive integer, as follows:

$${}_t p_x = \prod_{k=0}^{n-1} {}_{\frac{t}{n}} p_{x + \frac{kt}{n}}. \quad (21)$$

This motivates the first of two questions: what happens as $n \rightarrow \infty$? The second (related) question is: how can we express or represent events in the life history as a function of

passing time? We have a compact notation (${}_t p_x$, ${}_t q_x$ and so on) for the probabilities of events in the life history, but no such notation for the events themselves; generally we must express events somewhat clumsily in words. We consider these questions in turn in the next two sections.

4.2 The Product-integral Representation of a Survival Probability

From the heuristic ${}_h p_x \approx 1 - \mu_x h \approx \exp(-\mu_x h)$, for small h , we have the important product-integral representation as $n \rightarrow \infty$ and $1/n \rightarrow 0^+$:

$$\exp\left(-\int_0^t \mu_{x+s} ds\right) = {}_t p_x \quad (22)$$

$$= \lim_{n \rightarrow \infty} \prod_{k=0}^{n-1} {}_{\frac{k}{n}} p_{x+\frac{k}{n}} \quad (23)$$

$$\equiv \prod_{s \in (0, t]} (1 - \mu_{x+s} ds), \quad (24)$$

see Appendix 2 or, for example, Andersen et al. (1993). The product-integral has the same Π symbol as an ordinary product over a finite or countable number of terms, but is distinguished (here) by the presence of ds in the integrand and by the variable s ranging over an interval of the real line, $s \in (0, t]$. Then by differentiation of ${}_t p_x$, the density function of the random future lifetime T_x , denoted by $f_x(t)$, is:

$$f_x(t) = {}_t p_x \mu_{x+t} \quad (25)$$

$$= \exp\left(-\int_0^t \mu_{x+s} ds\right) \mu_{x+t} \quad (26)$$

$$= \prod_{s \in (0, t]} (1 - \mu_{x+s} ds) \mu_{x+t}. \quad (27)$$

Identities (271) and (27) are important in parametric mortality models, because they allow the likelihood to be specified entirely in terms of the hazard rate.

4.3 The Counting Process Representation of the Data

Suppose the i th individual has future lifetime T_0^i , a non-negative random variable. Define the process $N^i(x) = I_{\{T_0^i \leq x\}}$. This has value 0 as long as the i th individual is alive, and value 1 if they are dead (including at exact age T_0^i). Thus $N^i(x)$ ‘counts’ the number of deaths up to and including age x .

Associated with $N^i(x)$ is the indicator of survival, denoted by $Y^i(x)$ and defined as:

$$Y^i(x) = I_{\{\textit{i}th \textit{ individual alive at age } x^-\}} = I_{\{T_0^i \geq x\}}. \quad (28)$$

Thus $Y^i(x)$ is almost, but not quite, equal to $1 - N^i(x)$. Both have value 1 at exact age $x = T_0^i$, the age at death (see Table 4). $N^i(x)$ is right-continuous, while $Y^i(x)$ is left-continuous. This is for technical reasons when forming integrals.

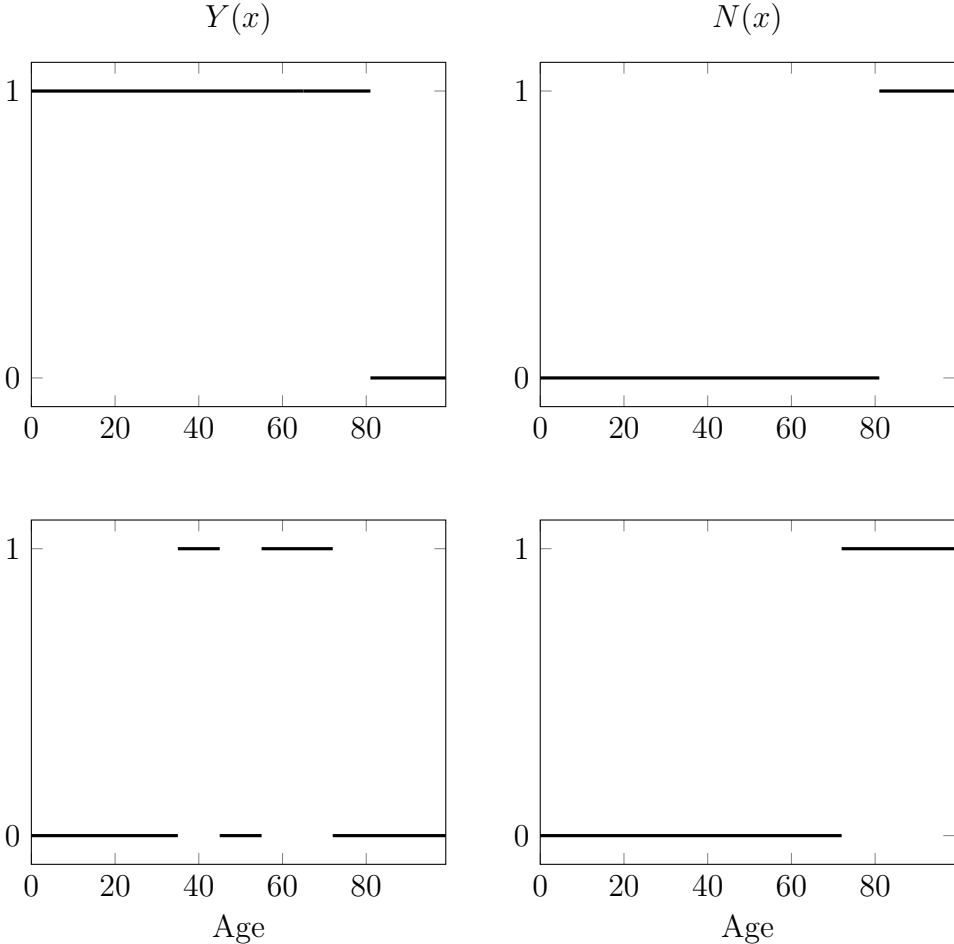


Figure 1: Sample counting-process representations of lifetimes. The first row is a life that enters observation at age 0 and is observed until dying at age 81. The second row is a life that enters observation at age 35 (left-truncation) and leaves at age 45 (right-censoring), enters observation again at age 55 (left-truncation) and is then observed until death at age 72.

Note that $N^i(x)$ represents exactly the same information as T_0^i , but in the form of a stochastic process instead of a random variable. It is purely descriptive, no probabilities or hazard rates appear in its definition, hence we refer to the counting process ‘representation’, not the counting process ‘model’.

Mention the word ‘process’ to an actuary under the age of about fifty, and it will trigger thoughts of Brownian motion, Itô calculus, stochastic integrals and option pricing. While important and necessary in its place, a counting processes carries none of that baggage². It really is nothing but a parsimonious way to represent an event happening at a random time, by means of zeros and ones. It must have been re-invented hundreds of times by computer programmers needing to represent events in binary.

Nevertheless, with $N^i(x)$ and $Y^i(x)$ representing the data, and μ_x as a model of the underlying ‘mechanism’ generating the data, we have the key to many problems of survival models.

4.4 The Multiplicative Model

Just above, we called μ_x the “... model of the underlying ‘mechanism’ generating the data ...”, referring of course to the heuristic, that the probability of death occurring between ages x and $x+h$, conditional on not having occurred beforehand, is approximately $\mu_x h$ (for small h). In fact, we make a small adjustment with a truly profound effect.

Define the stochastic hazard rate at age x to be the product $Y^i(x) \mu_x$, also called the Aalen multiplicative model (Andersen et al. 1993) (both names derive from the fact that the hazard rate is multiplied by a stochastic indicator). This represents a hazard rate tailored to the i th individual, that is automatically switched ‘on’ while they are alive and ‘off’ at any other time. Where before, we have had to qualify almost everything we said with the mantra “conditional on the life being alive at age x ” or the like, this is taken care of by the stochastic hazard rate. This explains the title ‘Dynamic Life History Models’ of this section; the hazard rates that govern the evolution of the life history are themselves stochastic and changed by events.

4.5 The Stochastic Probability Function: Back to Bernoulli

The probability function of the life history up to age x , from equation (27), in its product-integral form, can be denoted by $f_0^i(x)$ and written as:

$$f_0^i(x) = \prod_{s \in (0, x]} (1 - Y^i(s) \mu_s ds)^{1 - dN^i(s)} (Y^i(s) \mu_s)^{dN^i(s)}. \quad (29)$$

This says, heuristically, that times when the i th individual does not die ($dN^i(s) = 0$) contribute a survival probability $(1 - \mu_s ds)$ to the product, while the moment of death ($dN^i(s) = 1$) contributes the death probability $\mu_s ds$ (but by convention the ds is not displayed, as $f_0^i(s)$ is then a density function). (In this case the presence of $Y^i(s)$ makes no difference, because the exponents $1 - dN^i(s)$ and $dN^i(s)$ do the job by themselves, but we shall see why it is present in Section 4.6 below. It does matter, however, that $Y^i(x) = dN^i(x) = 1$ when x is the age at death.) Table 4 (‘Untruncated/Uncensored’)

²At least until the Central Limit Theorem is encountered, in which the limiting process is an Itô process, see Andersen et al. (1993).

Table 4: Contributions to the probability function of the infinitesimal Bernoulli trials (equation (29)) from elements of the observed life history, in the absence of left-truncation and right-censoring, and in their presence. Technical point: $N^i(s)$ has right-continuous sample paths and $Y^i(s)$ has left-continuous sample paths (Section 4.3), so at the time of an observed death $N^i(s) = Y^i(s) = 1$.

Period/Time	Observed	$Y^i(s)$	$N^i(s)$	$dN^i(s)$	$df_0^i(x)$
<u>Untruncated/Uncensored</u>					
Before death	Yes	1	0	0	$(1 - Y^i(s)) \mu_s ds$
Time of death	Yes	1	1	1	$Y^i(s) \mu_s ds$
After death	Yes	0	1	0	1
<u>Truncated/Censored</u>					
Before death	No	0	0	0	1
Before death	Yes	1	0	0	$(1 - Y^i(s)) \mu_s ds$
Time of death	No	0	1	1	1
Time of death	Yes	1	1	1	$Y^i(s) \mu_s ds$
After death	No	0	1	0	1
After death	Yes	0	1	0	1

shows the contributions to the likelihood (29) at different points in the observed life history.

In other words, at every time s when the i th individual is alive there is an infinitesimal Bernoulli trial with probability of death $Y^i(s) \mu_s ds$. See Gill (1994) on an infinite Bernoulli process. Some authors make product-integration the starting point of survival analysis (Cox & Oakes 1984, Lancaster 1990, Kalbfleisch & Prentice 2002). Actuaries are so strongly oriented towards binomial and Poisson models, however, that we have approached product-integration from there.

4.6 Left-truncation and Right-censoring

Left-truncation arises when the first part of a lifetime is unobserved. It is a fundamental characteristic of actuarial data, given that the vast majority of insured lives only become known to the insurer as adults. Right-censoring arises when the lifetime leaves observation before the event of interest (such as death) has occurred. There are many causes of right-censoring; Richards & Macdonald (2024) discuss a wide range of right-censoring events in the context of pensions and annuities.

Left-truncation and right-censoring can, in most cases, be allowed for very simply by adjusting the definition of the indicator $Y^i(s)$, as follows:

$$Y^i(s) = I_{\{\textit{i}th \textit{individual alive and under observation at age } s^-\}}. \tag{30}$$

With this change, everything said in Sections 4.1 to 4.5, including the important representation in equation (29), remains valid. Table 4 (‘Truncated/Censored’) shows the contributions to the likelihood (29) at different points in the observed and unobserved life

history, where $Y^i(s)$ indicates the presence of left-truncation and right-censoring by taking the value 0. Equation (29) as derived in Section 4.3 describes a purely mathematical probabilistic model. Substituting the indicator processes in equation (30) turns it into the basis of a statistical model involving data.

Figure 1 shows values of $Y^i(x)$ and $N^i(x)$ for two individuals. In the first row, we see observation of the complete lifetime T_0^i from birth to death, here at age 81. This illustrates the $Y^i(x)$ of equation (28). In the second row, one who enters observation at age 35 (left-truncation) and leaves without dying at age 45 (right-censoring), then re-enters observation at age 55 (left-truncation) and dies at age 72. This could happen if, for example, extracts from two different policy files are found during data cleaning to refer to the same individual. This illustrates the $Y^i(x)$ of equation (30).

The impact of this change in simplifying the mathematics is more than its apparent innocence would suggest. An example will be seen in Section 4.7.

4.7 Example: Individual versus Age-grouped Data for Multiple Lives Again

In Section 3.5 and Appendix 1 we showed that the likelihoods for individual and age-grouped data were the same, and pseudo-Poisson in form. The method was to derive contributions to the likelihood arising from the smallest possible ‘units’ of exposure to risk, namely the intersection of the i th individual’s lifetime and the k th rate interval. The total likelihood was then the product of all these ‘unit’ likelihoods over all individuals and all rate intervals. The proof in Appendix 1 is not technically difficult, but is burdened with the notation needed to define intervals and their intersections. By way of contrast we give below an alternative proof using the counting process representation of the data.

The contribution of the i th individual to the likelihood is:

$$L_i = \prod_{s \in (0, \infty]} (1 - Y^i(s) \mu_s ds)^{1 - dN^i(s)} (Y^i(s) \mu_s)^{dN^i(s)}. \quad (31)$$

Therefore the total likelihood is:

$$L = \prod_i L_i = \prod_i \prod_{s \in (0, \infty]} (1 - Y^i(s) \mu_s ds)^{1 - dN^i(s)} (Y^i(s) \mu_s)^{dN^i(s)}. \quad (32)$$

Split the age range into rate intervals Δ_k , noting that there is no contribution outside the age range $[r_0, r_K]$:

$$L = \prod_i \prod_{k=1}^K \prod_{s \in \Delta_k} (1 - Y^i(s) \mu_s ds)^{1 - dN^i(s)} (Y^i(s) \mu_s)^{dN^i(s)}. \quad (33)$$

Now change the order of the two outer products:

$$L = \prod_{k=1}^K \left[\prod_i \prod_{s \in \Delta_k} (1 - Y^i(s) \mu_s ds)^{1 - dN^i(s)} (Y^i(s) \mu_s)^{dN^i(s)} \right] = \prod_{k=1}^K L_k \quad (34)$$

noting that the terms in large brackets in (34) are the contributions from each rate interval³, which we denote by L_k .

The simplicity and directness of the proof above arises from the simplicity of the range of integration of the innermost integral, namely 0 to ∞ instead of an interval defined by the intersection of two other intervals, see Appendix 1. This reinforces an observation by Lidstone (1905) to the effect that it may be simpler to investigate what happens moment-by-moment, than over an extended interval:

“... it will be found that the formulae are in reality simplified through the absence of any distinction between the beginning and end of the momentarily intervals under consideration.” (Lidstone 1905)

Of course, the work in actually computing such complicated integrals is unchanged, but the greater ease of comprehension makes the task of the theorist (and the reader!) much easier.

4.8 A Classification of Mortality Models for Actuarial Use

Figure 2 illustrates how the models underlying survival analysis and occurrence-exposure rates used by actuaries all derive from Bernoulli trials over different time intervals.

- (a) The upper branch goes through the instantaneous Bernoulli trial with parameter $\mu_x dx$, which through product integration over the age interval Δ_i and simple aggregation over $M > 1$ individuals leads to the survival model based on individual lives.
- (b) Similarly, product-integration restricted to the rate interval Δ_k , and a constant hazard rate, leads to the pseudo-Poisson model.
- (c) The Poisson model also belongs to the upper branch, but would require a special (and unlikely) observational plan to ensure that E_x^c is deterministic and M is random. An example would be to replace each individual who dies with an identical individual until E_x^c reaches a pre-determined level (Scott 1982).
- (d) The lower branch goes through the Bernoulli trial with parameter q_x and leads to the binomial model with $M > 1$ individuals.

Note that both branches lose information about which individuals died in going from the individual model ($M = 1$) to the collective model ($M > 1$) but this will not matter for inference since the likelihood will be changed only by a factor not involving the parameter (for example, in the lower branch, a binomial coefficient). Only in the lower branch, however, is information also lost about the times of death, and therefore the true total of person-years lived. This is genuine loss of information which matters for inference, leading us to prefer any of the models in the upper branch over the binomial model.

³A.D.Wilkie, in written comments in the discussion of Richards (2008), advanced a similar argument based on “small age steps h ” and a quasi-indicator function similar to $Y^i(x)$; the argument given above is the same as in (Macdonald et al. 2018, Section 5.7).

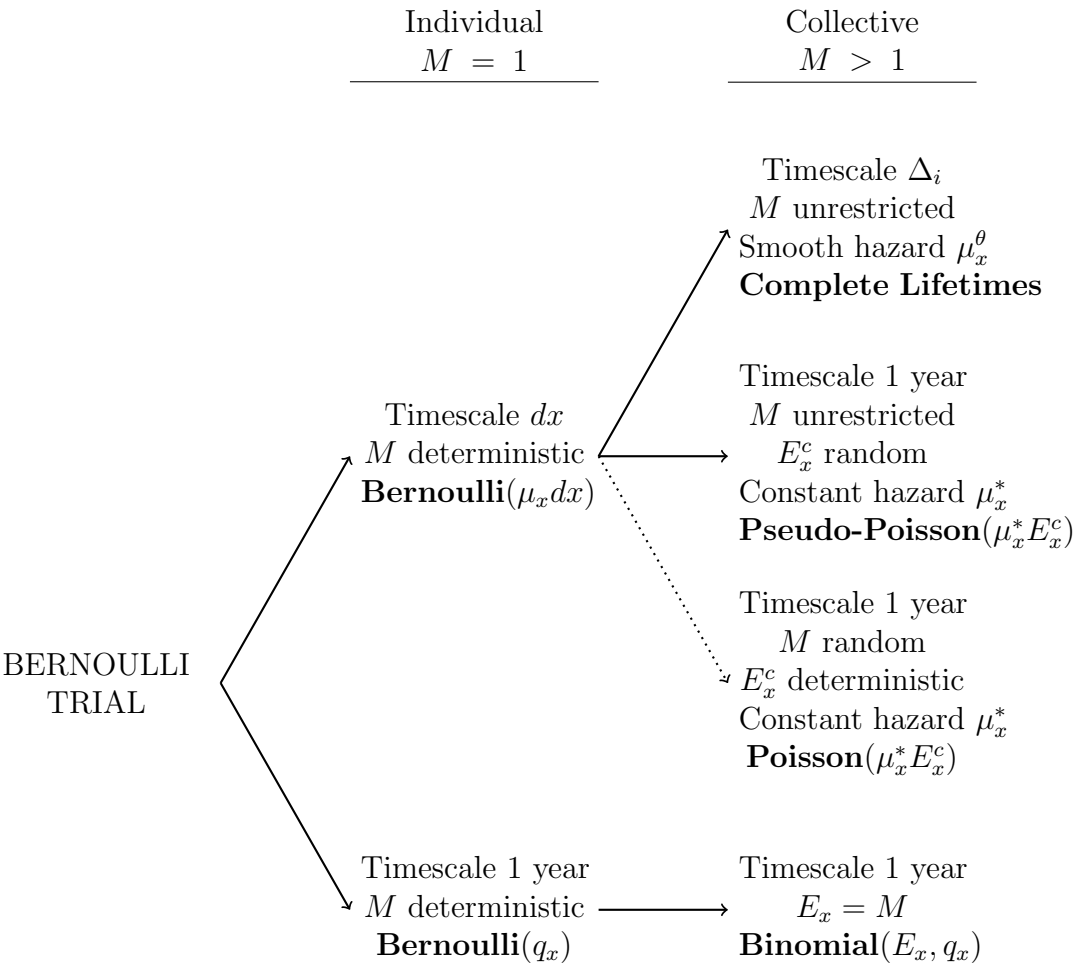


Figure 2: Family tree of models, showing the derivation of individual models, and collective models for individual lives and age-grouped data, (in **bold**) from the basic Bernoulli trial. M is the number of individuals observed, ‘ M unrestricted’ means that M can be either random or deterministic. The tree has two branches, one at the top leading to continuous-time models, including pseudo-Poisson and Poisson models, and one at the bottom leading to the discrete-time binomial model. The dotted arrow indicates that the Poisson model requires the imposition of an observational plan that ensures E_x^c is deterministic, which is unlikely to be realized in practice.

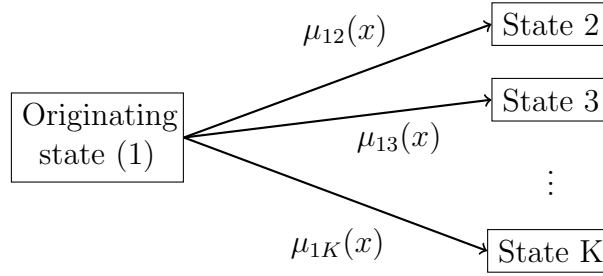


Figure 3: Multiple-decrement model. Note that states 2, 3, . . . , K are absorbing states with no transitions out once entered.

4.9 Extension to Multiple Decrements

Equation (29) is easily extended to multiple-state models, with a set of states labelled 1, 2, . . . , K. Omitting detailed definitions, process $Y_j^i(x)$ indicates observed presence in state j , and process $N_{jk}^i(x)$ ($j \neq k$) counts transitions from state j to state k , governed by stochastic intensity $Y_j^i(x) \mu_{jk}(x)$, and everything proceeds as before. We will not pursue this in its full generality, referring the reader to Macdonald et al. (2018) (elementary) or Andersen et al. (1993) (advanced), but the simplicity of this extension is an attractive feature of the counting process representation.

However we will sketch briefly the extension to multiple decrement models, since this has been a staple of actuarial textbooks. We have one originating state, labelled 1, in which all life histories begin, and $K - 1$ decrement states, labelled 2, 3, . . . , K. Transitions are possible from state 1 to any decrement state, governed by intensity $Y_1^i(x) \mu_{1k}(x)$ for the i th individual⁴. Figure 3 illustrates this model. Intensities out of a given state are additive, so exit from state 1 is represented by the counting process $N_{1\bullet}^i(x) = N_{12}^i(x) + \dots + N_{1K}^i(x)$, governed by the total intensity, denoted by $\mu_{1\bullet}(x)$, defined as $\mu_{1\bullet}(x) = \mu_{12}(x) + \dots + \mu_{1K}(x)$. The probability function of the life history can be expressed in two rules:

Rule 1: The time of exit of the i th individual from state 1 has probability function similar to equation (29):

$$f^i(x) = \prod_{s \in (0, x]} (1 - Y_1^i(s) \mu_{1\bullet}(s) ds)^{1 - dN_{1\bullet}^i(s)} (Y_1^i(s) \mu_{1\bullet}(s))^{dN_{1\bullet}^i(s)}. \quad (35)$$

Rule 2: Conditional on the i th individual exiting state 1 at age x , the probability that the state entered was k is $\mu_{1k}(x)/\mu_{1\bullet}(x)$, $k = 2, \dots, K$.

Therefore the model is specified completely by a product of infinitesimal Bernoulli trials as in Rule 1, and a simple ratio of intensities as in Rule 2.

If, however, the analyst begins by specifying a binomial-type model, for example based on a time unit of a year, then it is not easy to obtain a convincing representation of behaviour over shorter time periods, and there is certainly no unique solution. The

⁴We could rule all other transitions to be impossible by decree, or assume that all intensities $\mu_{jk}(x)$ with $j \neq 1$ are zero.

classical actuarial approach (see Neill (1977) or Bowers et al. (1997)) involves specifying a hypothetical model of each decrement acting alone, leading to a ‘gross’ hazard rate acting in the presence of the other decrements, and a ‘net’ hazard rate acting in their absence⁵. Except in some special cases, further progress is impossible unless ‘gross’ and ‘net’ hazard rates are assumed to be equal, but then since ‘gross’ hazard rates can be shown to be those of the Markov multiple-state model anyway, the modeller is drawn ineluctably towards that destination. Going into more detail would require too much notation, see (Macdonald et al. 2018, Chapter 16).

This illustrates vividly the contrast between the simplicity of the model specified at the ‘micro’ level, from which, by aggregation, behaviour at the ‘macro’ level can be deduced; and the perils of specifying the model at the ‘macro’ level, and then trying to disaggregate it to deduce or intuit behaviour on smaller scales.

4.10 Further Applications

We mentioned in Section 4.9 the extension of the counting process representation to multiple-state models, see Macdonald et al. (2018). We mention here other advances and applications based on counting processes, which can be found in the references below.

- (a) counting process compensators, martingales, stochastic integrals and central limit theorems (Andersen et al. 1993, Chapter III);
- (b) non-parametric estimates including the Nelson-Aalen and Kaplan-Meier estimates (Andersen et al. 1993, Chapters IV.1 and IV.3), (Kalbfleisch & Prentice 2002);
- (c) non-parametric kernel smoothing methods (Andersen et al. 1993, Chapter IV.2);
- (d) semi-parametric regression models including the Cox model and partial likelihoods (Andersen et al. 1993, Chapter 7), (Kalbfleisch & Prentice 2002);
- (e) log-rank comparison tests of survival models (Andersen et al. 1993, Chapter 5), (Kalbfleisch & Prentice 2002);
- (f) stochastic reserving models in life insurance Norberg (1991); and
- (g) stochastic models of surplus in life insurance, including Hattendorff’s theorem (Ramlau-Hansen (1988a,b), Norberg (1991)).

5. CONCLUSIONS

In search of sound foundations for mortality models, we began with binomial and Poisson models for grouped counts (Section 2), the basis of traditional graduations of mortality data for actuarial use. Both being models for ‘count’ data — the number of deaths during a time period, typically a year — they naturally invite questions about mortality at a smaller scale, over a fraction of the time unit. The Poisson model has an answer; its parameter is a hazard rate μ_x , assumed to be constant over the time period, and that defines mortality over short time intervals h as $h \rightarrow 0^+$. The binomial model has no such answer; it is up to the modeller to assume how mortality behaves over shrinking time intervals h as $h \rightarrow 0^+$. While this is unsatisfactory, the binomial

⁵Thanks to one of the more confusing legacies of actuarial nomenclature, ‘gross’ and ‘net’ hazard rates are traditionally called ‘dependent’ and ‘independent’ forces of decrement, which have absolutely no connection to statistical notions of dependence and independence.

model does decompose into a sum of Bernoulli trials, each representing the mortality of an individual over the time period.

Modelling the lifetime of a single individual (Section 3) gives the vital insight that the associated likelihood, as in equation (13), has the same form as a Poisson likelihood. Indeed we get the same Poisson-like likelihood if we model individual life histories, or group the data by age (Section 3.5), but in neither case do Poisson random variables feature as part of the model. That is, inference proceeds correctly as if the death counts we observe were Poisson random variables, but they are not. We suggest that all such mortality models based on age-grouped data — which includes most published models — should be called pseudo-Poisson models.

Finally, we identify the fundamental element of a mortality model as the infinitesimal Bernoulli trial; heuristically, an individual alive at age x will die in small time h with probability $h\mu_x$, or survive with probability $1 - h\mu_x$. To write down probabilities of events over extended time intervals, we need to know how to aggregate such trials, and that requires three ideas new to most actuaries:

- (a) the product-integral (Section 4.2 and Appendix 2), as the method of aggregating probabilities of infinitesimal Bernoulli trials over extended intervals;
- (b) counting processes (Section 4.3), giving us the natural notation to describe the events in a life history; and
- (c) the stochastic hazard rate $Y^i(x)\mu_x$ tailored to the life history of the i th individual including left-truncation and right-censoring (Section 4.6).

In conclusion, equation (29), with $Y^i(x)$ as in equation (30), is the ‘atom’ of a ‘continuous-time’ survival model, with the qualities listed in Section 1.5 and reproduced below.

- (a) It is irreducible, in the sense that it is composed of (infinitesimal) Bernoulli trials.
- (b) It is based on behaviour at the ‘micro’ time scale.
- (c) It is based on individual lives.
- (d) Aggregated, over time and over individuals, it explains the Poisson-like nature of likelihoods, therefore estimation based on the collective at the ‘macro’ time scale.
- (e) It allows for left-truncation and right-censoring.
- (f) It is easily extended to multiple-decrement and multiple-state models.

ACKNOWLEDGEMENTS

We are grateful to Mr Gavin Ritchie for comments on a draft of this paper.

COMPETING INTERESTS

None.

APPENDIX 1

THE LIKELIHOODS FOR INDIVIDUAL AND AGE-GROUPED DATA

We show that the likelihood obtained by modelling individual lifetimes (see Section 3.5) is the same as that obtained from age-grouped data for each rate interval. To do so we use the factorization in Section 3.3. Note three points before we begin:

- (a) The equality of likelihoods means that inference has the same results using either approach, although each may have advantages for other reasons.
- (b) We show equality assuming an arbitrary form of hazard rate μ_x , making no parametric assumptions. This includes as a special case the assumption of piecewise-constant hazard rates, constant on each rate interval, usually made in conjunction with the assumption that total deaths D_x in each rate interval are Poisson random variables.
- (c) The result here, which resides in the equality of expressions in equations (40) and (41), is quite detailed, and follows by selecting elements of three different partitions of the age range in precise ways. This contrasts with Section 4.7, where the same result follows as an easy consequence of the definitions in the counting process representation of the data.

Suppose we have M lives, not all identical, the i th individual being observed between ages x_i and $x_i + v_i$ and the random variable d_i indicating death or censoring at age $x_i + v_i$. Define $y_i = x_i + v_i$, and let Δ_i be the interval $(x_i, y_i]$ on which the i th individual is observed.

We wish to introduce a set of rate intervals and write down contributions to the likelihood for the i th individual over those rate intervals that intersect Δ_i . We need some detailed definitions, which we introduce in three steps.

Step 1: Intervals: Let the sequence of ages $r_0 < r_1 < \dots < r_K$, with $r_0 \leq \min_i x_i$ and $r_K \geq \max_i y_i$, define the rate intervals $\Delta_k = (r_{k-1}, r_k]$ ($k = 1, \dots, K$). Let $w_0 < w_1 < \dots < w_J$ be the sequence formed by the (ordered) union of the three sequences $x_1, \dots, x_M, y_1, \dots, y_M$ and r_0, \dots, r_K and define $\Delta_j = (w_{j-1}, w_j]$ ($j = 1, \dots, J$). Hence we have rate intervals Δ_k , and the i th individual exposed to risk on interval Δ_i , and the intervals Δ_j are formed from the intersections of all Δ_i and Δ_k . Figure 4 illustrates these three sequences of intervals.

Step 2: Indicators: For ($j = 1, \dots, J$), define the sequences of indicators:

$$e_{i,j} = \begin{cases} 1 & \text{if } \Delta_j \subseteq \Delta_i \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

and:

$$d_{i,j} = \begin{cases} d_i & \text{if } w_j = y_i \\ 0 & \text{otherwise.} \end{cases} \quad (37)$$

Therefore $e_{i,j}$ indicates that the i th individual was exposed to risk during $(w_{j-1}, w_j]$, and $d_{i,j}$ indicates death or censoring of the i th individual at age w_j .

Step 3: Likelihoods: We can now replace likelihood (13) with the following:

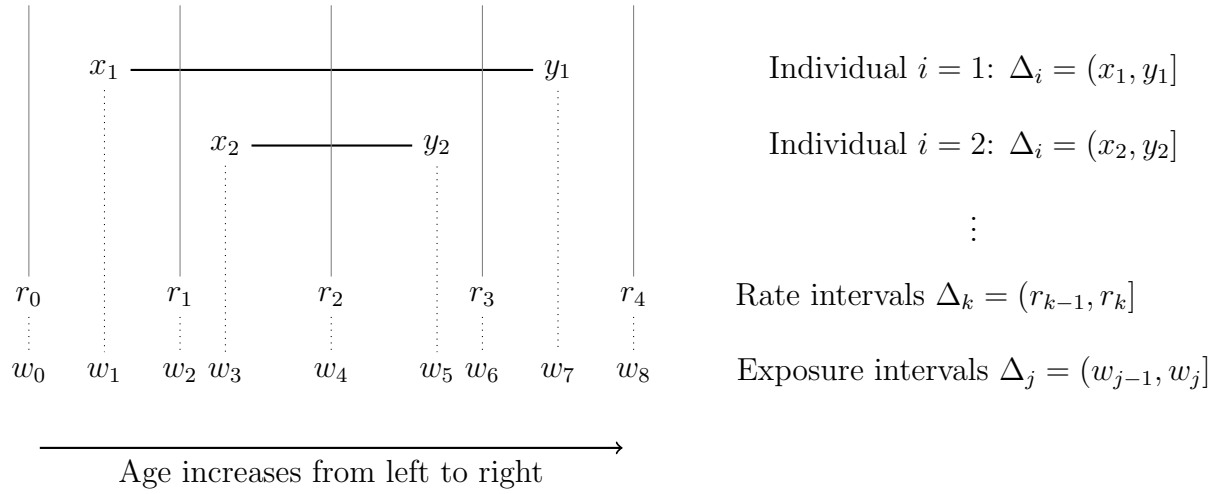


Figure 4: Rate intervals, individual observations and exposure intervals. Rate-interval boundaries, $\{r_k\}$, are set by the analyst, and here are not necessarily integers, nor evenly spaced. The data are paired ages of the start and end of individual observations, (x_i, y_i) . The set of exposure-interval boundaries, $\{w_j\}$, is defined as the ordered union of $\{r_k\}$, $\{x_i\}$ and $\{y_i\}$. One consequence is that each exposure interval Δ_j is always completely contained within a corresponding rate interval Δ_k .

$$L_i = \prod_{j=1}^J \left[\exp \left(- \int_{\Delta_j} \mu_s ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}} \quad (38)$$

and collect together all those intervals Δ_j that are part of the rate interval Δ_k :

$$L_i = \prod_{k=1}^K \prod_{\Delta_j \subseteq \Delta_k} \left[\exp \left(- \int_{\Delta_j} \mu_s ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}}. \quad (39)$$

The payoff from all this careful defining of points and intervals comes when we form the total likelihood over all M individuals, denoted by L :

$$L = \prod_{i=1}^M L_i = \prod_{i=1}^M \prod_{k=1}^K \prod_{\Delta_j \subseteq \Delta_k} \left[\exp \left(- \int_{\Delta_j} \mu_s ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}}. \quad (40)$$

Then by reversing the order of the two outer products, we can collect together contributions to each rate interval instead of contributions to each lifetime:

$$L = \prod_{k=1}^K \prod_{i=1}^M \prod_{\Delta_j \subseteq \Delta_k} \left[\exp \left(- \int_{\Delta_j} \mu_s ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}} = \prod_{k=1}^K L_k \quad (41)$$

where we define L_k as:

$$L_k = \prod_{i=1}^M \prod_{\Delta_j \subseteq \Delta_k} \left[\exp \left(- \int_{\Delta_j} \mu_s ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}}, \quad (42)$$

which we recognize as the total contribution for rate interval Δ_k . The desired equality is shown above, namely $\prod_i L_i = \prod_k L_k$. We note that the intricacy of the definitions and the argument (including the presence of two sets of indicators $d_{i,j}$ and $e_{i,j}$) stems from the need to handle both points and intervals of time, in different combinations. This need is largely abolished, as far as algebra is concerned, by the definition of the process $Y^i(x)$ in Section 4.3.

APPENDIX 2

THE PRODUCT-INTEGRAL

The ordinary integral is familiar to actuaries, the product-integral less so. However, every time an actuary multiplies survival probabilities of the form ${}_t p_x$, she uses a product-integral. It is clear that she uses an ordinary integral, and moreover uses its additive property, since:

$$\begin{aligned} {}_{t+s} p_x &= \exp \left(- \int_0^{t+s} \mu_{x+r} dr \right) \\ &= \exp \left(- \int_0^t \mu_{x+r} dr - \int_t^{t+s} \mu_{x+r} dr \right) \\ &= \exp \left(- \int_0^t \mu_{x+r} dr \right) \exp \left(- \int_0^s \mu_{x+t+r} dr \right) \\ &= {}_t p_x {}_s p_{x+t}. \end{aligned} \quad (43)$$

This suggests the exponential function as a link between functions with additive and multiplicative properties, and indeed it is. Start with the following identity, proved in most courses on real analysis (see (Hardy 1992, pp. 410–411)):

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \right)^n = e. \quad (44)$$

More generally, assuming we may exchange logarithms and limits and then taking just the first-order term of the Taylor expansion $\log(1+s) = s - s^2/2 + s^3/3 - \dots$ (convergent on $-1 < s \leq 1$):

$$\log \lim_{n \rightarrow \infty} \left(1 + \frac{s}{n} \right)^n \approx \lim_{n \rightarrow \infty} n \frac{s}{n} = s \quad (45)$$

implying (44) and more. This is homogeneous, in the sense that $n \times s/n = \sum_1^n s/n$ is a sum of n equal summands. Suppose we have a well-behaved function $f(s)$ on an interval $(a, b]$. Partition the interval into n equal sub-intervals denoted by $\Delta_1 = [a, a+h), \dots, \Delta_n = (b-h, b]$ where $h = (b-a)/n$, and let $f(s_k)$ be the function value at an arbitrarily chosen $s_k \in \Delta_k$ ($k = 1, 2, \dots, n$). Then by the same reasoning:

$$\log \lim_{n \rightarrow \infty} \prod_{k=1}^n (1 + f(s_k) h) \approx \lim_{n \rightarrow \infty} \sum_{k=1}^n f(s_k) h = \int_a^b f(s) ds. \quad (46)$$

It only remains to replace the interval length h with the more general ds in the limit and exponentiate both sides, giving us the important representation:

$$\prod_{s \in (a,b]} (1 + f(s) ds) = \exp \left(\int_a^b f(s) ds \right). \quad (47)$$

Choose the function $f(s) = -\mu_{x+s}$ on the interval $(0, t]$ as in a survival probability, and we have the product-integral representation of the familiar identity:

$$\prod_{s \in (0,t]} (1 - \mu_{x+s} ds) = \exp \left(- \int_0^t \mu_{x+s} ds \right) = {}_t p_x \quad (48)$$

which is equation (24).

The above is intuitive and heuristic; for a rigorous account see Gill & Johansen (1990) and references therein.

References

- Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993), Statistical Models Based on Counting Processes, Springer, New York.
- Batten, R. W. (1978), Mortality Table Construction, Prentice-Hall.
- Benjamin, B. & Pollard, J. H. (1980), The Analysis of Mortality and Other Actuarial Statistics, Heinemann, London.
- Bowers, N., Gerber, H., Hickman, J., Jones, D. & Nesbitt, C. (1997), Actuarial Mathematics, 2nd edition, Society of Actuaries, Schaumburg IL.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. & Balevich, I. (2009), ‘A quantitative comparison of stochastic mortality models using data from England and Wales and the United States’, North American Actuarial Journal **13**, 1–35.
- Cox, D. R. (1972), ‘Regression models and life-tables’, Journal of the Royal Statistical Society, Series B (Methodological) **34**, 187–220.
- Cox, D. R. & Oakes, D. (1984), The Analysis of Survival Data, Chapman-Hall, London.
- Currie, I. D. (2016), ‘On fitting generalized linear and non-linear models of mortality’, Scandinavian Actuarial Journal **2016**, 356–383.
- Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2020), Actuarial Mathematics for Life Contingent Risks, International Series on Actuarial Science, third edn, Cambridge University Press.

- Forfar, D. O., McCutcheon, J. J. & Wilkie, A. D. (1988), ‘On graduation by mathematical formula’, Journal of the Institute of Actuaries **115**, 1–149.
- Gill, R. D. (1994), Lectures on survival analysis, in P. Bernard, ed., ‘Lectures on Probability Theory. Ecole d’Eté de Probabilités de Saint-Fleur XXII–1992’, Springer-Verlag, Berlin, Heidelberg.
- Gill, R. D. & Johansen, S. (1990), ‘A survey of product-integration with a view toward application in survival analysis’, Annals of Statistics **18**, 1501–1555.
- Hardy, G. H. (1992), A Course in Pure Mathematics, Cambridge University Press.
- Kalbfleisch, J. D. & Prentice, R. L. (2002), The Statistical Analysis of Failure Time Data (second edition), John Wiley & Sons, Hoboken NJ.
- Lancaster, T. (1990), The Econometric Analysis of Transition Data, Cambridge University Press, Cambridge.
- Lidstone, G. J. (1905), ‘Changes to pure premium policy-values consequent upon variations in the rate of interest or the rate of mortality, or upon the introduction of the rate of discontinuance (with discussion)’, Journal of the Institute of Actuaries **39**, 209–252.
- Macdonald, A. S. (1996), ‘An actuarial survey of statistical models for decrement and transition data I: Multiple state, Poisson and binomial models’, British Actuarial Journal **2**, 129–155.
- Macdonald, A. S., Richards, S. J. & Currie, I. D. (2018), Modelling Mortality with Actuarial Applications, International Series on Actuarial Science, Cambridge University Press.
- Neill, A. (1977), Life Contingencies, Heinemann, London.
- Nelder, J. A. & Wedderburn, R. W. M. (1972), ‘Generalized linear models’, Journal of the Royal Statistical Society. Series A (General) **135**, 370–384.
- Norberg, R. (1991), ‘Reserves in life and pension insurance’, Scandinavian Actuarial Journal **1991**, 3–24.
- R Core Team (2021), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Ramlau-Hansen, H. (1988a), ‘The emergence of profit in life insurance’, Insurance: Mathematics and Economics **7**, 225–236.
- Ramlau-Hansen, H. (1988b), ‘Hattendorff’s theorem: A Markov chain and counting process approach’, Scandinavian Actuarial Journal **1988**, 143–156.
- Richards, S. J. (2008), ‘Applying survival models to pensioner mortality data (with discussion)’, British Actuarial Journal **14**, 257–326.

Richards, S. J. & Macdonald, A. S. (2024), ‘On contemporary mortality models for actuarial use I: Practice’, Working Paper .

Scott, W. F. (1982), ‘Some applications of the Poisson distribution in mortality studies’, Transactions of the Faculty of Actuaries **38**, 255–263.

Sverdrup, E. (1965), ‘Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between states of health’, Skandinavisk Aktuaritidskrift **48**, 184–211.

Waters, H. R. (1984), ‘An approach to the study of multiple state models’, Journal of the Institute of Actuaries **111**, 363–374.

Wintringham, T. & Blashford-Snell, J. N. (1973), Weapons and Tactics, Penguin Books Ltd., Harmondsworth.